



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Genetic basis of hypersusceptibility to protease inhibitors and low replicative capacity of human immunodeficiency virus type 1 strains in primary infection**

**Citation for published version:**

Leigh Brown, AJ, Frost, SDW, Good, B, Daar, ES, Simon, V, Markowitz, M, Collier, AC, Connick, E, Conway, B, Margolick, JB, Routy, JP, Corbeil, J, Hellmann, NS, Richman, DD & Little, SJ 2004, 'Genetic basis of hypersusceptibility to protease inhibitors and low replicative capacity of human immunodeficiency virus type 1 strains in primary infection', *Journal of Virology*, vol. 78, no. 5, pp. 2242-2246.  
<https://doi.org/10.1128/JVI.78.5.2242-2246.2004>

**Digital Object Identifier (DOI):**

[10.1128/JVI.78.5.2242-2246.2004](https://doi.org/10.1128/JVI.78.5.2242-2246.2004)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Journal of Virology

**Publisher Rights Statement:**

Freely available via Pub Med.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Genetic Basis of Hypersusceptibility to Protease Inhibitors and Low Replicative Capacity of Human Immunodeficiency Virus Type 1 Strains in Primary Infection

Andrew J. Leigh Brown,<sup>1\*</sup> Simon D. W. Frost,<sup>2</sup> Benjamin Good,<sup>2</sup> Eric S. Daar,<sup>3</sup> Viviana Simon,<sup>4</sup> Martin Markowitz,<sup>4</sup> Ann C. Collier,<sup>5</sup> Elizabeth Connick,<sup>6</sup> Brian Conway,<sup>7</sup> Joseph B. Margolick,<sup>8</sup> Jean-Pierre Routy,<sup>9</sup> Jacques Corbeil,<sup>2</sup> Nicholas S. Hellmann,<sup>10</sup> Douglas D. Richman,<sup>2,11</sup> and Susan J. Little<sup>2</sup>

*University of Edinburgh, Edinburgh, Scotland<sup>1</sup>; University of California, San Diego,<sup>2</sup> and VA San Diego Healthcare System,<sup>11</sup> La Jolla, Harbor-UCLA, Los Angeles,<sup>3</sup> and ViroLogic Inc., South San Francisco,<sup>10</sup> California; Aaron Diamond AIDS Research Center, New York, New York<sup>4</sup>; University of Washington, Seattle, Washington<sup>5</sup>; University of Colorado Health Sciences Center, Denver, Colorado<sup>6</sup>; University of British Columbia, Vancouver,<sup>7</sup> and McGill University Health Centre, Montreal,<sup>9</sup> Canada; and Johns Hopkins University, Baltimore, Maryland<sup>8</sup>*

Received 27 May 2003/Accepted 17 October 2003

The initial virus strains from as many as 12% of individuals with primary human immunodeficiency virus (HIV) infection have a 50% inhibitory concentration  $\leq 0.4$ -fold that of HIV type 1<sub>NL4-3</sub> (HIV-1<sub>NL4-3</sub>) to ritonavir (hypersusceptibility [HS]). There is also substantial variation in replicative capacity (RC) or an in vitro assay of the contributions of protease (PR) and reverse transcriptase to viral fitness. In chronically infected antiretrovirally treated patients, amprenavir HS has been associated with the mutation N88S in PR, but this mutation is not seen in untreated patients. In this study, virus strains from 182 cases of primary HIV infection were analyzed, and a highly significant association between HS and low RC ( $\leq 10\%$  that of HIV-1<sub>NL4-3</sub>) was observed ( $P < 10^{-6}$ ). Multivariate analysis was used to determine the genotypic basis of ritonavir HS, analyzing all polymorphic amino acid sites and insertions from p7gag through PR. Decision tree models developed on the entire Gag-plus-PR data set and on PR alone gave overall correct classifications of 73 and 72%, respectively, on cross-validation. They were also able to predict low RC, with sensitivities of 69 and 62% and specificities of 84 and 70%, respectively. The analysis shows that ritonavir HS in untreated primary HIV infection is not associated with single mutations but with combinations of amino acids at polymorphic sites and that the same genotypes which confer HS to PR inhibitors confer low RC. This supports the view that variation in PR function is directly responsible for variation in fitness among strains in primary infection.

Variation in fitness and other phenotypic traits in human immunodeficiency virus type 1 (HIV-1) is a natural consequence of the mechanism of retroviral replication and the error rate associated with reverse transcriptase (9). Advances in methodologies for assaying susceptibility to antiretroviral (ARV) drugs (13) have led to improved methods, which have also reduced the errors associated with estimating replicative capacity (RC), a component of viral fitness. Use of this single-cycle assay has revealed a surprising range of values in virus from ARV-naïve patients, from as low as 6% that of the reference strain, HIV-1<sub>NL4-3</sub>, to as much as 50% higher (T. Wrin, A. Gamarnik, N. Whitehurst, J. Beachaine, J. M. Whitcomb, N. S. Hellmann, and C. J. Petropoulos, abstr. 24, 5th Int. Workshop HIV Drug Resistance and Treatment Strategies, 4–8 June 2001). The genetic basis for this variation remains unclear. It has been known for some time that fitness of viral strains with reduced susceptibility to ARVs, especially to viral protease (PR) inhibitors (PIs), can be substantially reduced (3, 10–12). However, those strains carried mutations which are not seen in untreated individuals, and therefore, a different mechanism must be responsible for low RC in wild-type virus.

Hypersusceptibility (HS) to ARV drugs, defined here as susceptibility  $\leq 0.4$ -fold that of HIV-1<sub>NL4-3</sub>, is another phenotype that was first described in the context of patients who had been treated with a failing ARV regimen; HS to nonnucleoside reverse transcriptase inhibitors has recently been shown to be clinically significant (4). In one study, among those who had acquired a virus strain resistant to nelfinavir,  $>6\%$  were found to be HS to amprenavir (18). In this situation, amprenavir HS was shown to be specifically associated with mutations at amino acid 88 in PR, particularly N88S (18). However, HS has also been described in a subset of individuals who have never received therapy (Wrin et al., Abstr. 5th Int. Workshop on HIV Drug Resistance), in whom this mutation is absent. We show here that, among patients with primary HIV infection, low RC and HS to PIs are directly related, and we obtain a single decision tree model for the genetic bases of both.

## MATERIALS AND METHODS

**Patients.** The subjects were a subset of those described by Little et al. (7), selected on the basis of availability of sequence data for both p7/p6gag and PR and not having  $>10$ -fold reduced susceptibility to any PI. They were recruited between April 1997 and May 2000 at clinics in 10 North American cities and were predominantly men who reported a history of sex with men. None of the subjects had received  $>7$  days of prior ARV therapy before study entry and analysis of ARV susceptibility. ARV susceptibility was determined using HIV PhenoSense. The genotypes of the p7 and p6 domains of gag and the PR domain of pol were obtained using ABI automated DNA sequencing.

\* Corresponding author. Mailing address: ICAPB, University of Edinburgh, West Mains Rd., Edinburgh EH9 3JT, Scotland. Phone: 44-131-650-5523. Fax: 44-131-650-6564. E-mail: A.Leigh-Brown@ed.ac.uk.

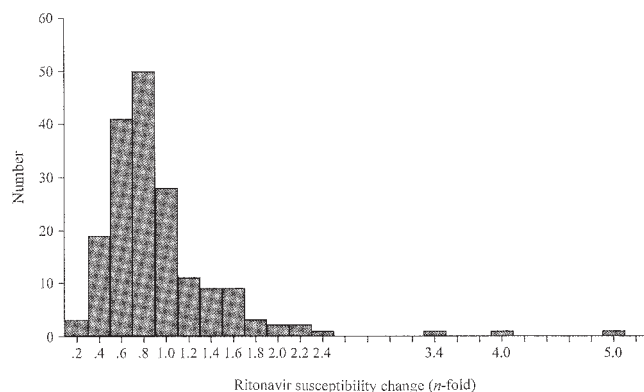


FIG. 1. Distribution of ritonavir susceptibilities in 182 primary-infection patients. Cases of transmitted PI-resistant virus (7) ( $>10$ -fold less than the susceptibility of HIV-1<sub>NL4-3</sub>) were not included in the analysis.

**Data.** All amino acid sites where the most common amino acid was present at a frequency  $\leq 98\%$  were included in the analysis. Mutations at each site were analyzed using single-letter codes.

**Techniques.** A variety of analytical methods were investigated, including stepwise logistic regression (SPSS version 10.1) and CART (S-Plus version 6.0) and the related informatics-based methods of decision trees, PART rules, and support vector machines implemented in the Weka package (17). CART trees and C4.5 decision trees produced similar results; the results from the C4.5 decision trees and support vector machines are given here.

The C4.5 decision trees were generated using a cost-sensitive classifier. A range of cost values and tip (leaf) sizes were explored for both the ritonavir and amprenavir analyses, and models were tested by 90-10 cross-validation. In this process, a model is repeatedly generated based on a 90% sample of cases chosen at random, and its prediction is tested on the remaining 10%. The tests were run 10 times on independent samples of the data to give the quoted sensitivities and specificities.

**Nucleotide sequence accession numbers.** The nucleotide sequences have been deposited in GenBank under accession numbers AY518941 to AY519122.

## RESULTS

**Frequency of HS to PIs in primary HIV infection.** The distribution of ritonavir susceptibilities among the 182 cases for which complete sequence (p7gag-PR) and phenotype data were available is shown in Fig. 1, excluding cases of transmitted PI resistance ( $>10$ -fold higher than that of HIV-1<sub>NL4-3</sub>) (7). The mean susceptibility for ritonavir was 0.9-fold that of HIV-1<sub>NL4-3</sub> (standard deviation, 0.58). Using the 0.4-fold cutoff, 22 (12.1%) cases were HS for ritonavir and 21 (11.5%) were HS for indinavir. In contrast, only 6 cases were HS for nelfinavir (3.3%), while 38 (20.9%) were HS for amprenavir. For saquinavir, 28 cases were HS (15.9%), and lopinavir was not included in these assays. Ritonavir seemed to be broadly representative of PI HS, as the 22 ritonavir HS strains included all 6 nelfinavir HS strains, 15 of 21 indinavir HS strains, 19 of 28 saquinavir HS strains, and 18 of 38 amprenavir HS strains. Thus, the major distinction among PIs, which is between nelfinavir and amprenavir, is explained by different mean values (Fig. 1). The mean susceptibility for nelfinavir is 1.3 times that for HIV-1<sub>NL4-3</sub>, so the curve is right shifted relative to ritonavir, while the mean for amprenavir is 0.77 times that for HIV-1<sub>NL4-3</sub>, so the distribution is left shifted relative to ritonavir, with many more strains therefore falling below the 0.4-fold threshold. In most cases where a strain was HS for one drug

and not for another, the susceptibility change values were within 0.1-fold (e.g., 0.4-fold [HS] for ritonavir but 0.5-fold for saquinavir). This difference is below the reproducibility of the assay, so many such cases could be considered to have the same drug susceptibility. Although errors are inevitable following the imposition of a threshold, most of the variation in susceptibility lies well above this range, and attempts to analyze susceptibility as a continuous variable did not generate models which explained variation at the lower end of the range (data not shown).

**Correlation between PI susceptibility and RC.** There was strong correlation between PI susceptibility and RC among these primary HIV strains (Fig. 2) (Pearson correlation coefficient, 0.5;  $P < 10^{-10}$ ). The mean RC was 43% of that of HIV-1<sub>NL4-3</sub> (median, 41%), and the relationship was better fitted by the quadratic curve shown in Fig. 2 (analysis of variance,  $F_2 = 43$ ) than by either a straight line or logistic ( $F_1 = 14$  for both).

The most striking feature of Fig. 2 is the close correlation at the bottom end of the range for both variables, to the extent that of the 13 strains classified as having low RC (i.e.,  $\leq 0.1$ -fold the RC of HIV-1<sub>NL4-3</sub>), 9 were also classified as HS to ritonavir (exact  $P$ ,  $< 10^{-6}$ ).

**Polymorphism in gag-PR.** The fact that there is a high level of polymorphism in PR in untreated patients has been known for some time (1, 5). Defining a polymorphic site with the criterion that the most common amino acid has a frequency  $\leq 98\%$ , 20 of the 99 amino acid sites in PR were polymorphic in this data set. In addition, of the 70 amino acid sites available for analysis from p7 and p6, 55 (79%) were polymorphic by the same criterion, and the region included four polymorphic insertions at gag454, gag460, gag478, and gag483 (numbering according to HIV-1<sub>LAI</sub>, clone HXB2R). This high level of variation is all the more remarkable because in the p6 region

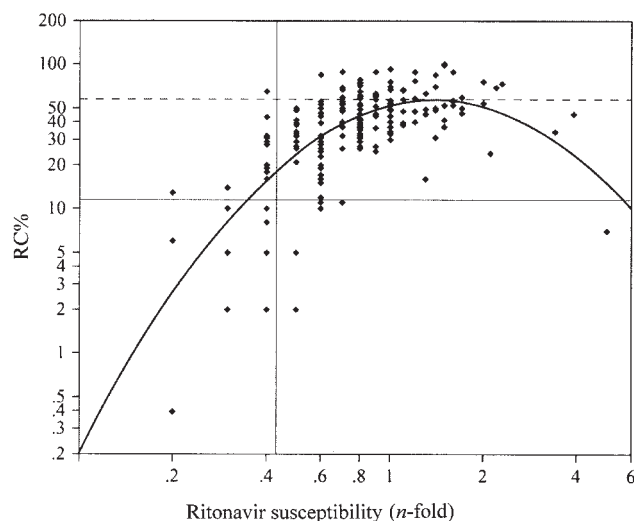


FIG. 2. Relationship between RC and ritonavir susceptibility. Both variables are plotted on  $\log_{10}$  scales. The continuous curved line represents the fitted quadratic function, which reaches a maximum at an RC value of 56% (dashed horizontal line). Points to the left of the vertical line are cases classified as HS to ritonavir (susceptibility,  $\leq 0.4$ -fold that of HIV-1<sub>NL4-3</sub>). The points below the solid horizontal line are cases classified as low RC ( $\leq 0.1$ -fold the RC of HIV-1<sub>NL4-3</sub>).

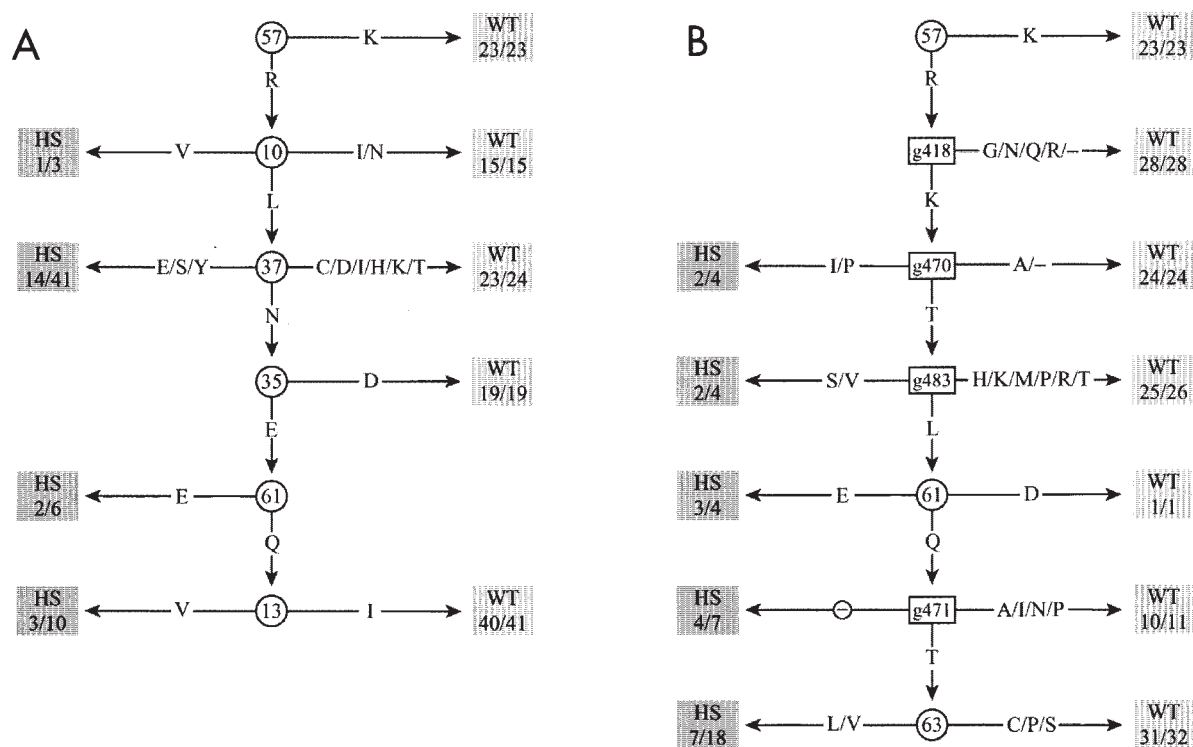


FIG. 3. Decision tree model describing the genetic basis for ritonavir HS in primary HIV strains. WT, wild type. (A) Based on PR alone. A cost value of 7.2 and a leaf size of 8 were used to obtain this model (see the text). (B) Based on gag plus PR. Amino acid sites in gag are preceded by g; -, amino acid deletion at the site. A cost value of 5 and a leaf size of 7 were used to obtain this model. The models are interpreted by checking the amino acid sites listed and following the prediction shown. Thus (for both), (i) if amino acid 57 is R, then information from amino acid 10 is used, while if 57 is K, then the wild type is predicted (in this data set, all 23 with this amino acid are wild type); (ii) in panel A, if amino acid 10 is L, position 37 is examined, and if 10 is I or N then the wild type is predicted (of 15 in the data set with the genotype 57K plus 10I/N, again, all are wild type).

the *pol* reading frame (−1 with respect to that of *gag*) overlaps with that of the *gag* polyprotein for 56 amino acids.

**Genetic basis of HS.** Initial studies were carried out using logistic regression with a modified data set of PR sites alone in which sites with more than one mutant amino acid were represented by a series of “dummy” binary variables, each corresponding to a single allele. A highly significant ( $P < 10^{-7}$ ) model incorporating five amino acid sites was identified (amino acids 12, 33, 37, 45, and 63). At amino acid 37, two amino acids (E and Y), and at amino acid 63, three amino acids (L, V, and Q) were independently associated with HS. However, this model had a sensitivity (rate of correct prediction of HS) of only 37.5%.

Structured models, such as those used in classification trees (2, 15), can be more powerful than simple logistic models where there are complex relationships between parameters. In addition, the generation of many additional variables in binary models results in overfitting. In contrast, machine learning methods, such as decision trees (16, 17), are very flexible and naturally accept multiple classifications for each variable. In addition, the models generated are explicit and can be compared with other available information. We also investigated another machine learning approach, support vector machines, but they do not readily permit explicit interpretation of the sites used in partitioning the data. We used two sequence data sets from the same samples: the first was based on PR alone

and was similar to that used in the logistic regression (above) but with multiple classifications of mutant amino acids. The second data set comprised the first set plus the 55 polymorphic amino acid sites and four polymorphic insertions in p7gag plus p6gag.

Given the frequency of HS in the data set, it was necessary to use cost-sensitive classifiers; otherwise, a misleadingly high overall prediction success rate (76%) could be obtained by misclassifying all HS cases as wild type, despite giving a sensitivity of 0%. Cost values varying between 6 and 8 for the PR data set and between 4 and 6 for the Gag-plus-PR data set were explored to find the model giving maximal sensitivity and specificity. Trees were pruned to improve generality, and the models were tested using 90-10 cross-validation (see Materials and Methods). The overall correct classifications for the optimal PR-based model (Fig. 3A) was 72%; that for the gag-plus-PR model (Fig. 3B) was 73%. For PR sites alone, the sensitivity (correct prediction of HS) was 73% and the specificity (correct prediction of wild type) was 68% (Table 1). For the model based on Gag-plus-PR data, the sensitivity was lower (59%) but the specificity was higher (75%). This represents almost a sixfold enrichment relative to the frequency of HS in the data set.

Surprisingly, despite the inclusion of all the PR sites in both data sets, the best model obtained from the full data set included only two of the sites in the PR model (57 and 61) and



TABLE 1. Ritonavir HS model performance on 10-fold cross-validation

	Model performance (%)			
	Decision tree		SVM <sup>a</sup>	
	PR alone	gag + PR	PR alone	gag + PR
Sensitivity	73	59	73	75
Specificity	68	75	72	68

<sup>a</sup> Support vector machine.

did not improve on it in terms of the accuracy of prediction of HS strains. To attempt to explain this, pairwise association tests were run between the four gag sites and the remaining four in the PR model. Of the 16 tests performed, one ( $37 \times \text{g418}$ ) was significant ( $P = 0.04$ ; Fisher's exact test) but lost significance when multiple tests were corrected for, leaving one ( $10 \times \text{g471}$ ;  $P = 0.0053$ ; Fisher's exact test) which remained significant when corrected. However, the association did not involve the amino acid deletion at g471, which is a predictor of HS, but rather involved the rare amino acids g471A/I/N/P, which were jointly associated with PR10I more frequently than expected.

**Prediction of HS to other PIs.** The ritonavir-derived model in Fig. 3A also predicts HS to the other PIs tolerably well (Table 2), with the exception of amprenavir. The sensitivities (the true-positive rates) for saquinavir and indinavir were  $>60\%$  (cf. 73% for ritonavir), while for nelfinavir it was 100%, indicating that all six nelfinavir HS cases were correctly identified. The lower sensitivity for amprenavir is not unexpected, as amprenavir has a left-shifted susceptibility distribution and nearly twice as many HS cases.

**HS and low RC.** We have already shown that in this data set there was a strong association between the phenotypes of low RC and HS (Fig. 2). Comparison of the predictions of HS and of low RC for each of the two data sets reveals that the ritonavir HS PR-based model had a sensitivity of 62% for low RC and a specificity of 70%, a ninefold enrichment for this phenotype. The gag-PR based model was actually slightly better (69 and 84%, respectively). We conclude that the genotypic bases of low RC and HS in strains from primary HIV infection are essentially the same. One possible reason for the association could be that it is a consequence of the assay, with strains with low RC inevitably being more readily inhibited. However, this would apply to all drugs, not just PIs, and there is no general increase in susceptibility to other ARVs in these strains (data not shown). These results suggest that the alterations in PR function associated with HS are responsible for the low RCs of these strains.

## DISCUSSION

We have analyzed the genetic bases of HS and low RC in strains from individuals with primary HIV infection who were not infected with drug-resistant strains, in terms of the genetic variation conferred by mutations at polymorphic amino acid sites. In previous work, it was shown that such mutations can be associated with variation in susceptibility to nonnucleoside reverse transcriptase inhibitors which is observed in primary HIV infection (6). This conclusion was tested in that study by

in vitro mutagenesis, which confirmed the in vivo observations. In the present study, however, HS could not be explained by just one or two polymorphic sites. Stepwise binary logistic regression on variable amino acids in PR identified a highly significant model incorporating amino acid sites 12, 33, 37, 45, and 63; however, it had a sensitivity of only 36%. The presence at several polymorphic amino acid sites of multiple mutant amino acids raises two possible approaches. Either all non-wild-type amino acids are pooled as "mutant," as in previous studies (6, 14), or in order to distinguish their individual contributions (in case different mutant amino acids have effects in different directions), it is possible to consider each amino acid mutation at a site as a separate variable. However, the additional variables that this introduces can result in overfitting of the model so that the results are not generalizable to other data sets. In addition, this treatment of different amino acid variants is unlikely to reflect the way in which these mutations interact at a phenotypic level.

To explicitly incorporate multiple amino acids at the same site, and to permit analysis of more complex models, a decision tree approach was used. In order to test the relevance of sites in gagp6/p7, the data were analyzed with and without the presence of gag sequences. Surprisingly, it was found that inclusion of the gag region in the data set did not improve the performance of the model. Both data sets yielded models that were  $\sim 72\%$  correct overall on cross-validation, and the Gag-plus-PR model had lower sensitivity on cross-validation than the PR-based model (Gag plus PR, 59%; PR alone, 73%).

One possible explanation for the lack of improvement in the model with the additional data from gag is that there were tight nonrandom associations between mutant amino acids at variable sites in gag and in PR. We therefore performed an analysis of nonrandom association among amino acids at the sites involved in the two models. This failed to identify any associations except one between amino acids 10 in PR and 471 in gag. However, given the level of variability in PR in this data set, the addition of the gag sites may not have provided any further information in the classification of the strains. The 20 amino acid sites included in the PR data set generated 161 distinct PR amino acid sequences among these 182 strains. This suggests that the PR-based model correctly identified the mechanistic basis of HS, while the additional sites from gag merely provided other material to classify the strains.

The extent of functional variation in PR, as defined by assays of susceptibility to PIs, is substantial. In this data set, from which transmitted resistant strains ( $>10$ -fold reduction in susceptibility) have been excluded, a 25-fold range in susceptibility to ritonavir was observed. A large range in RC was also observed. The relationship between these two is closest at the lowest values for each: five of the seven with susceptibilities of

TABLE 2. Prediction of HS for other PIs by the ritonavir HS PR-based decision tree model

	Model performance (%)				
	Ritonavir	Indinavir	Saquinavir	Nelfinavir	Amprenavir
Sensitivity	73	62	62	100	40
Specificity	68	71	73	70	69

<0.4 had low RCs. At higher values, the two are not the same—only 4 out of the observed 13 HS cases with susceptibility values of 0.4 also had low RCs. One possible explanation for very low average RC values is the inclusion of a mixture of viable and completely inviable virus, e.g., with termination codons in coding sequences, in these samples. However, a frequency reached by a termination codon sufficient to cause the effect would be detectable in the consensus genotype, in which mixtures of >25% would become detectable as ambiguities. The frequency of ambiguities in these samples was much lower than required in this scenario (data not shown).

Despite the low measured RCs, it is clear that all of these virus strains had successfully established infections within an average of ~70 days prior to being sampled (7). Clearly, they possess the basic requirements to establish an infection in a naïve host, and yet many appear to be a long way from the “optimal” RC value (the median RC for drug-susceptible strains is ~70% of that of HIV-1<sub>NL4-3</sub>). In an earlier study,  $R_0$ , the basic reproductive rate, in acute HIV infection was estimated to be ~20, with a range from 7 to 34 (8).  $R_0$  has to be >1 to permit an infection to be self-sustaining. Assuming the laboratory-measured RC is an additive component of absolute fitness in vivo, this would suggest that the low-RC strains could have an  $R_0$  of ~2, which could still permit them to establish an active infection. Thus, we conclude that the variation in RC is associated with the extreme end of a continuous spectrum of variation in fitness, to which genotypic variation in the PR sequence contributes heavily.

#### ACKNOWLEDGMENTS

We are grateful to Gregor M. Cardué for help with the figures.

Financial support was received from the Universitywide AIDS Research Program, University of California, grant numbers PH97-SD-201 and PH97-CS-202; Center for AIDS Research grant number AI 36214; General Clinical Research Center National Center for Research Resources grant M01-RR00425 and M01-RR00102; grant numbers AI 41534, AI 43638, AI 47033, AI 47745, and TW00767 (Fogarty Center) from the National Institutes of Health; the Research Center for AIDS and HIV Infection of the San Diego Veterans Affairs Medical Center; the Canadian Institutes of Health Research and Fonds de Recherche en Santé du Québec; and an unrestricted donation from Roche Molecular Systems.

#### REFERENCES

1. Barrie, K. A., E. E. Perez, S. L. Lamers, W. G. Farmerie, B. M. Dunn, J. W. Sleasman, and M. M. Goodenow. 1996. Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. *Virology* **219**: 407–416.
2. Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. Classification and regression trees. Wadsworth International, Belmont, Calif.
3. Croteau, G., L. Doyon, D. Thibeault, G. McKercher, L. Pilote, and D. Lamarre. 1997. Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. *J. Virol.* **71**:1089–1096.
4. Haubrich, R. H., C. A. Kemper, N. S. Hellmann, P. H. Keiser, M. D. Witt, D. N. Forthal, J. Leedom, M. Leibowitz, J. M. Whitcomb, D. Richman, and J. A. McCutchan. 2002. The clinical relevance of non-nucleoside reverse transcriptase inhibitor hypersusceptibility: a prospective cohort analysis. *AIDS* **16**:F33–F40.
5. Kozal, M. J., N. Shah, N. Shen, R. Yang, R. Fucini, T. C. Merigan, D. D. Richman, D. Morris, E. Hubbell, M. Chee, and T. R. Gingeras. 1996. Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nat. Med.* **2**:753–759.
6. Leigh Brown, A. J., H. M. Precious, J. M. Whitcomb, J. K. Wong, M. Quigg, W. Huang, E. Daar, R. T. D'Aquila, P. H. Keiser, E. Connick, N. S. Hellmann, C. J. Petropoulos, D. D. Richman, and S. J. Little. 2000. Reduced susceptibility of HIV-1 from patients with primary HIV infection to non-nucleoside reverse transcriptase inhibitors is associated with variation at novel amino acid sites. *J. Virol.* **74**:10269–10273.
7. Little, S. J., S. Holte, J. P. Routy, E. S. Daar, M. Markowitz, A. C. Collier, R. A. Koup, J. W. Mellors, E. Connick, B. Conway, M. Kilby, L. Wang, J. M. Whitcomb, N. S. Hellmann, and D. D. Richman. 2002. Antiretroviral-drug resistance among patients recently infected with HIV. *N. Engl. J. Med.* **347**:385–394.
8. Little, S. J., A. R. McLean, C. A. Spina, D. D. Richman, and D. V. Havlir. 1999. Viral dynamics of acute HIV-1 infection. *J. Exp. Med.* **190**:841–850.
9. Mansky, L. M., and H. M. Temin. 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* **69**:5087–5094.
10. Martinez-Picado, J., A. V. Savara, L. Shi, L. Sutton, and R. T. D'Aquila. 2000. Fitness of human immunodeficiency virus type 1 protease inhibitor-selected single mutants. *Virology* **275**:318–322.
11. Martinez-Picado, J., A. V. Savara, L. Sutton, and R. T. D'Aquila. 1999. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J. Virol.* **73**:3744–3752.
12. Nijhuis, M., R. Schuurman, D. de Jong, J. Erickson, E. Gustchina, J. Albert, P. Schipper, S. Gulnick, and C. A. B. Boucher. 1999. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* **13**:2349–2359.
13. Petropoulos, C. J., N. T. Parkin, K. L. Limoli, Y. S. Lie, T. Wrin, W. Huang, H. Tian, D. A. Smith, G. A. Winslow, D. J. Capon, and J. M. Whitcomb. 2000. A novel phenotypic drug susceptibility assay for HIV-1. *Antimicrob. Agents Chemother.* **44**:920–928.
14. Precious, H. M., H. F. Günthard, J. K. Wong, R. T. D'Aquila, V. A. Johnson, D. R. Kuritzkes, D. D. Richman, and A. J. Leigh Brown. 2000. Multiple regression analysis of sequence variants in HIV-1 reverse transcriptase at onset of antiretroviral therapy can predict long term virological response. *AIDS* **14**:31–36.
15. Quigg, M., S. D. W. Frost, S. McDonagh, S. M. Burns, D. Clutterbuck, A. McMillan, C. L. S. Leen, and A. J. Leigh Brown. 2002. Association of antiretroviral resistance genotypes with response to therapy—comparison of three models. *Antivir. Ther.* **7**:151–157.
16. Quinlan, J. R. 1993. C4.5: programs for machine learning. Morgan Kaufman, San Francisco, Calif.
17. Witten, I. H., and E. Frank. 2000. Data mining. Morgan Kaufman, San Francisco, Calif.
18. Ziermann, R., K. Limoli, K. Das, E. Arnold, C. J. Petropoulos, and N. T. Parkin. 2000. A mutation in human immunodeficiency virus type 1 protease, N88S, that causes in vitro hypersensitivity to amprenavir. *J. Virol.* **74**:4414–4419.